Aug 2020

2020

CAPSTONE PROJECT

Regions to open Lounge/Casino in the city of Kuala Lumpur, Malaysia

By: Ashutosh Agrawal

Contents

Introduction

With the growing demand of the people to spend some quality time by going in some chill-out zones to make fun, the Casinos/Lounges are the best place for them. For the investors, this is a big chance to open such places to gather more crowd. It is not as simple as it looks to open such places at any region/place as they need to take care for many feasibilities and the increasing competition day by day. Under such conditions, tt is quite complicated for them to make wise choice.

These properties let the investors to earn consistently and permanently for the longer duration and here arises our business problem.

Business Problem

The major concern for the investors to open Casino/Fun Zone in which locality/place in Kuala Lumpur, Malaysia to gather more and more national/international crowd as there are many foreigners in the city as well. This project will help the property investors in the selection of best locality/place to open Casinos or some Chill-out zones.

To help the investors in making the wise choice, we will use the machine learning algorithms and data science practices to suggest the investors the best locations to open Casino/Chill-out zones.

Target Audience

The target audience includes the investors and the property dealers those who can start their business or make passive income by investing wisely in such zones. The idea here is to suggest them the best places to minimize the competition and let them earn fair profits for which they are actively looking for.

Required Data

The data required for this project comprises of the below section:

- **List of neighbourhoods for the city of Kuala Lumpur, Malaysia**
- **Coordinates of the neighbourhoods**
- **Venue categories in the neighbourhood**

Sources

All the three sections of the data mentioned above will be fulfilled from the different methodologies.

The wikipedia link "https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur" consists of the 71 neighbourhoods for the city Kuala Lumpur, Malaysia. We can use various python libraries like BeautifulSoup to extract the data using web scraping and Pandas dataframe to put the resultant data into a ready to use format. Pandas dataframe would consist of a list of all the neighbourhoods for the city Kuala Lumpur, Malaysia.

For every neighbourhoods of Kuala Lumpur, we will need the latitude & longitude coordinates to fetch the venues further. We can use Geocoder library in the python to fetch the coordinates for all the 71 neighbourhoods extracted above.

After getting all the above required data, we can finally use the FourSquare APIs to get all the Venues and the Venues Category in the city of Kuala Lumpur, Malaysia. FourSquare APIs is a very easy and efficient way to get the venues data and is used widely to fetch such kind of information.

Once we get all this data, we can finally apply clustering on the neighbourhoods for the Venues Category that consists of "Casino"/"Lounge" to find the locations where we can suggest to build new Casinos where there are no or less competition comparatively. In the next section of report we will discuss complete methodology to implement this project.

Methodology Followed

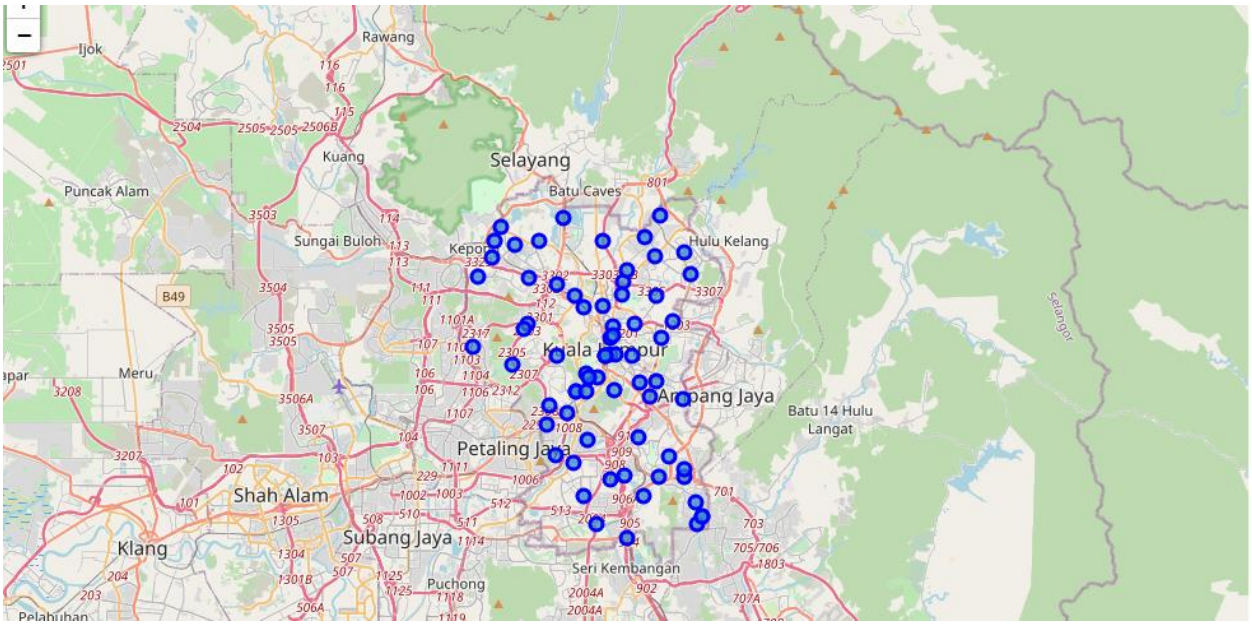
The methodologies consists of three sections:

- Fetching neighbourhoods of the city.
- Fetching Venues in the neighbourhoods.
- Clustering venues to get insights.

Fetching neighbourhoods of the city

In order to determine the best location for the Casino/Lounge in the Kuala Lumpur, Malaysia, we first required the neighbourhoods area data for the Kuala Lumpur, Malaysia. We got this data using the wiki page link "https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur". Using web page scraping technique by python using BeautifulSoup package, we were able to get the required neighbourhoods. But getting the neighbourhoods was not sufficient as we required to get the geographical coordinates as well to plot the neighbourhoods on the map of Kuala Lumpur, Malaysia. We used the Geocoder package of the python to fetch the latitude-longitude coordinates of the neighbourhoods fetched from the Wikipedia page.

Finally, we used the Folium package of the python to plot these neighbourhoods on the map of Kuala Lumpur, Malaysia city.



Fetching Venues in the neighbourhoods

As we have already got the neighbourhoods of the city in the previous step, we now required the venues and the venues category details of these neighbourhoods. We used the FourSquare APIs to get these details and to analyze it futher. FourSquare is a wonderful platform which provides different APIs to explore a city in regards with lat-long details. We used the API and get the venues details for the neighbourhoods details we had earlier. Finally, created a dataframe using Pandas library in python. The dataframe is such that it contains the details for the two kinds of venues which is Casino and the Lounge (as required in the project).

Clustering Venues to get insights

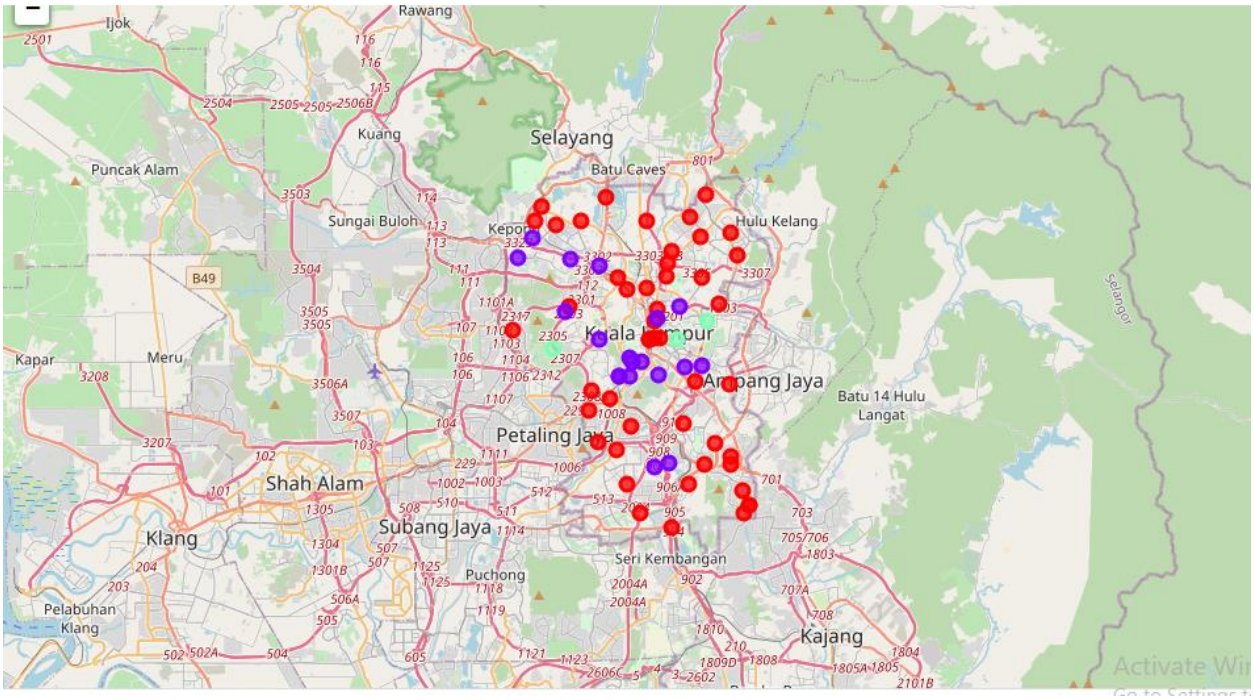
After we have got the required data from the previous steps, we then applied the k-means unsupervised machine learning algorithm on the "Casino/Lounge" data to put them in the clusters. K-means algorithm works on the centroid detection where it puts every data points in one of the cluster based on the occurrence and keeping the centroid as small as possible. In our project, we divided 3 clusters based on the occurrence of the keywords "Casino" and the "Lounge". Once its done we had the required analysis and the prediction ready to be used by the investors and the property dealers.

Results

We have divided the clusters into three parts : Cluster-0, Cluster-1 and the Cluster-2. These clusters have been divided on the basis of the occurrence of the venues category 'Casino' and the 'Lounge'. Below are our predictions after clustering analysis:

- Cluster-0 has only one Casino and no Lounges.
- Cluster-1 has no Casino but fair number of Lounges.
- Cluster-2 has comparatively high number of Lounges.

Resultant Map



The map consists of all the three clusters mentioned above. The clusters are segregated by the three colors: Red, Purple and the Mint green.

Discussion

As clearly observed from the map of the results section that the Cluster-0 is having no Lounges and a only Casino present. Cluster-1 has no Casinos but there is a fair competition when it comes to Lounges. On the other hand, Cluster-2 is having high number of Lounges present but with no Casino. We highly recommend property developers to open Casino in any of the areas in the Cluster-1 and the Cluster-2 but if they want to go for the Lounges then they should open it in the areas of the Cluster-0 as there is no Lounge present there.

Conclusion

This project covers all the aspects of the data science. We have gone through the various phases of the data science practices like

- Problem identification.
- Specifications of the data required.
- Data refinement and preparation.
- Visualizations.
- Machine learning
- Solution and the recommendations.

Ofcourse, the solution of the business problem stated in the introduction section is the areas under Cluster-0. Cluster-0 in the results section is the only area where property dealers and the investors can invest for no competition in the Lounges and for the Casinos there will be no competition if they go with Cluster-2 or Cluster-3 locations.

References

Kuala Lumpur, Malaysia neighbourhoods data source:

https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur

FourSquare Doc Page:

https://developer.foursquare.com/docs